

CLAIMS

What is claimed is:

- 1 1. A method for improving reliability and availability of a load balanced server
2 comprising the steps of:
 - 3 monitoring the server's performance;
 - 4 detecting when the server's performance is worse than a failover threshold; and
 - 5 sending a message to one or more clients indicating that said one or more clients
6 should failover to an alternate server.
- 1 2. The method of Claim 1, wherein the server is an AAA server and the one or more
2 clients are AAA clients.
- 1 3. The method of Claim 1, wherein the step of sending a message comprises sending an
2 ICMP Echo message.
- 1 4. The method of Claim 1, wherein the step of monitoring the server's performance
2 comprises measuring one or more parameters from the group consisting of server related
3 parameters, system related parameters, and availability of services on the server.
4
- 1 5. The method of Claim 4, wherein the server related parameters comprise a currently
2 available number of threads and a maximum number of available threads.
- 1 6. The method of Claim 4, wherein the system related parameters comprise CPU usage
2 percentage, memory usage percentage, network availability, and number of processes
3 running.
- 1 7. The method of Claim 4, wherein the services of which the availability is checked on
2 the server comprise mandatory services and dependant services.

1 8. The method of Claim 1, further comprising the step of determining the one or more
2 clients to which to send the message based on a predefined list of clients.

1 9. The method of Claim 1, further comprising the step of determining the one or more
2 clients to which to send the message based on a network device group.

1 10. The method of Claim 1, further comprising the step of determining the one or more
2 clients to which to send the message based on network topology.

1 11. The method of Claim 1, further comprising the step of determining the alternate
2 server based on a list configured on each of said one or more clients.

1 12. The method of Claim 1, wherein the message that is sent to said one or more clients
2 comprises a list of one or more alternate servers to which said one or more clients can
3 failover.

1 13. The method of Claim 1, further comprising the step of checking authority of a
2 message sent between a sender and a receiver by comparing a first hashed value, produced by
3 the sender and sent with the message, with a second hashed value produced by the receiver.

4 14. The method of Claim 13, further comprising the step of producing the first hashed
5 value and the second hashed value using a one-way hash algorithm with a shared secret as a
6 key and a combination of the server's IP address and the client's IP address as input.

1 15. The method of Claim 13, further comprising the step of producing the first hashed
2 value and the second hashed value using a one-way hash algorithm with a combination of a
3 shared secret, the server's IP address, and the client's IP address as input.

1 16. The method of Claim 1, further comprising the step of connecting with a second
2 client.

1 17. The method of Claim 16, further comprising the step of initiating the step of
2 connecting based on a request from the second client.

1 18. The method of Claim 17, further comprising the step of initiating the step of
2 connecting based on a timeout mechanism configured on the second client.

1 19. The method of Claim 16, further comprising the step of initiating the step of
2 connecting based on a request by the server.

1 20. The method of Claim 19, further comprising the step of initiating the step of
2 connecting based on the server's performance being better than a connection threshold.

1 21. The method of Claim 20, wherein the step of initiating comprises the step of
2 comparing the connection threshold with a function relating one or more parameters from the
3 group consisting of server related parameters, system related parameters, and availability of
4 services on the server.

1 22. The method of Claim 21, wherein the server related parameters comprise a currently
2 available number of threads and a maximum number of available threads.

1 23. The method of Claim 21, wherein the system related parameters comprise CPU usage
2 percentage, memory usage percentage, and number of processes running.

1 24. The method of Claim 21, wherein the services of which the availability is checked on
2 the server comprise services mandatory for correct functioning of the server and services
3 needed for logging on the server.

1 25. The method of Claim 16, wherein said one or more clients comprise multiple clients,
2 the method further comprises the steps of:

3 connecting a first set of one or more clients at a first time, wherein said first set of one
4 or more clients comprises one or more clients from said multiple clients; and
5 connecting a second set of one or more clients at a second time, wherein said first
6 time is different than said second time, and said second set of one or more
7 clients comprises one or more clients from said multiple clients.

1 26. The method of Claim 1, wherein said one or more clients comprise all clients
2 connected to said server.

1 27. The method of Claim 1, wherein said one or more clients comprise a proper subset of
2 all clients connected to said server.

1 28. The method of Claim 1, further comprising the steps of:
2 disconnecting a first set of one or more clients, wherein said first set of one or more
3 clients comprise one or more clients from said one or more clients; and
4 connecting a second set of one or more clients, wherein the second set of one or more
5 clients comprise one or more clients from said first set of one or more clients.

1 29. The method of Claim 28, wherein the step of connecting comprises the steps of:
2 connecting each client of said second set of one or more clients at a different time;
3 and
4 initiating the step of connecting each client based on a timeout mechanism configured
5 on each client.

1 30. The method of Claim 28, further comprising the step of initiating the step of
2 connecting based on the server's performance being better than a connection threshold,
3 wherein the server's performance is measured as a function relating one or more parameters
4 from the group consisting of server related parameters, system related parameters, and
5 availability of services on the server.

1 31. The method of Claim 28, wherein said second set of one or more clients comprises
2 multiple clients, and the step of connecting a second set of one or more clients comprises the
3 steps of:

4 connecting a third set of one or more clients at a first time, wherein said third set of
5 one or more clients comprises one or more clients from said multiple clients;
6 and

7 connecting a fourth set of one or more clients at a second time, wherein said first time
8 is different than said second time, and said second set of one or more clients
9 comprises one or more clients from said multiple clients.

1 32. The method of Claim 28, wherein said second set of one or more clients comprises all
2 of said one or more clients.

1 33. A computer-readable medium carrying one or more sequences of instructions for
2 improving reliability and availability of a load balanced server, which instructions, when
3 executed by one or more processors, cause the one or more processors to carry out the steps
4 of:

5 monitoring the server's performance;
6 detecting when the server's performance is worse than a failover threshold; and
7 sending a message to one or more clients indicating that said one or more clients
8 should failover to an alternate server.

1 34. An apparatus for improving reliability and availability of a load balanced server,
2 comprising:

3 means for monitoring the server's performance;
4 means for detecting when the server's performance is worse than a failover threshold;
5 and
6 means for sending a message to one or more clients indicating that said one or more
7 clients should failover to an alternate server.

1 35. An apparatus for improving reliability and availability of a load balanced server,
2 comprising:

3 a network interface that is coupled to the data network for receiving one or more packet
4 flows therefrom;
5 a processor;
6 one or more stored sequences of instructions which, when executed by the processor, cause
7 the processor to carry out the steps of:
8 monitoring the server's performance;
9 detecting when the server's performance is worse than a failover threshold; and
10 sending a message to one or more clients indicating that said one or more clients
11 should failover to an alternate server.